We claim:

1. A computer implemented method of generating context vectors representing information elements for retrieval of the information elements or records containing the information elements, the method comprising:

assigning a context vector to each of a plurality of information elements;

initializing the context vectors such that the context vectors are substantially orthogonal to each other in a vector space;

determining proximal co-occurences of the information elements; and

adjusting the context vectors based on the proximal co-occurrences of the information elements, such that the information elements that frequently proximally co-occur have context vectors with similar orientations in the vector space.

2. The method of claim 1, wherein initializating the context vectors further comprises:

assigning vector components to the vectors using zero-mean, unit-variance Gaussian random number generation.

3. The method of claim 1, wherein a target context vector is a context vector assigned to a target information element, and a neighbor context vector is a context vector assigned to an information element that proximally co-occurs with the target context vector, and wherein adjusting the context vectors comprises:

for each target context vector to be adjusted:

determining an error vector between a target context vector and each neighbor context vector;

updating the target context vector as a function of the error vectors.

4. The method of claim 3, wherein updating the target context vector as a function of the error vectors comprises:

determining a correction vector from the error vectors, where the correction vector is:

$$C_j = \sum_{i}^{WS} (\|E_{ij}\| - \alpha_{ij})\hat{E}_{ij}$$

where:

E_{ij} is the error vector $E_{ij} = N_{ij} - T_j$ between the neighbor context vector N_{ij} and a target context vector T_j;

WS is a window size containing the target context vector and the neighbor context vectors; and

α is a proximity constraint; and

updating the target context vector as

$$T_j^{NEW} = T_j^{OLD} + \frac{\gamma}{F_j} \sum_{i=1}^{F_j} C_j - M$$

where:

γ is a step size;

F_j is the total number of occurrences of information element j; and

M is a mean context vector for all unique context vectors.

5. The method of claim 1 wherein a target context vector is a context vector assigned to a target information element, and a neighbor context vector is a context vector assigned to an information element that proximally co-occurs with the target context vector, and wherein adjusting the context vectors comprises:

determining a weighted sum vector of neighbor context vectors of a target context vector;

applying the weight sum vector to the target context vector.

6. The method of claim 5, further comprising:

determining the weighted sum vector according to the equation:

$$W_j = \sum_{i}^{WS} \frac{G(i)}{D_j} N_{ij}$$

where:

W is the weighted sum vector;

N_{ij} is the neighbor context vector to target context vector T_j;

G(i) is a Gaussian weight for the neighbor context vector i; and

D_j is the number of documents that contain target information element j; and

applying the weighted sum vector to the target context vector according to the equation:

$$T_j^{NEW} = T_j^{OLD} + W_j$$

where:

T_j^{NEW} is the updated target context vector; and

T_j^{OLD} is the un-updated target context vector.

7. The method of claim 1 wherein a target context vector is a context vector assigned to a target information element, and a neighbor context vector is a context vector assigned to an information element that proximally co-occurs with the target context vector, and wherein adjusting the context vectors comprises:

determining a weighted sum vector of neighbor context vectors of a target context vector according to the equation

$$W_j = \sum_{i}^{WS} \frac{G(i)}{D_j} N_{ij}$$

where:

W is the weighted sum vector;

G(i) is a Gaussian weight for the neighbor context vector i; and

D_j is the number of documents that contain target information element j;

determining an error vector from the weighted sum vector and the target context vector:

$$E_j = W_j - T_j$$

where

E is the error vector;

T is target context vector;

determining a correction vector C from the error vectors of the neighbor context vectors:

$$C_j = \sum_{i=1}^{F_j} E_j$$

applying the correct vector to the target vector:

$$T_j^{NEW} = T_j^{OLD} + \gamma C_j - M$$

where:

T_j^{NEW} is the updated target context vector; and

T_j^{OLD} is the un-updated target context vector.

γ is a step size; and

M is a mean context vector for all unique context vectors.

8. The method of claim 1, wherein the vector space is defined by a plurality of axes, and the context vector include vector components corresponding to the respective axes, and wherein the axes individually do not have specific semantic associations.

9. The method of claim 1, wherein a target context vector is a context vector assigned to a target information element, and a neighbor context vector is a context vector assigned to an information element that proximally co-occurs with the target context vector, and wherein adjusting the context vectors comprises:

adjusting the target context vector as a function of the relative importance of at least one of either the target information element or the neighbor context vector with respect to the plurality of information elements.

10. The method of claim 1, wherein a target context vector is a context vector assigned to a target information element, and a neighbor context vector is a context vector assigned to an information element that proximally co-occurs with the target context vector, and wherein adjusting the context vectors comprises:

adjusting the target context vector as a function of the frequency of occurrence of the target information element; the frequency of occurrence of the neighbor information element; a total number of records containing the target information element, and a total number of record containing the neighbor information element.

11. The method of claim 1, wherein a target context vector is a context vector assigned to a target information element, and a neighbor context vector is a context vector assigned to an information element that proximally co-occurs with the target context vector, and wherein adjusting the context vectors further comprises:

adjusting the target context vector as a function of:

a distance of the neighbor information element from the target information element, so that neighbor information elements that are closer to the target information element cause the target context vector to be adjusted to be closer to the neighbor context vector; and

a frequency of occurrence of the neighbor information element in records containing the information elements, such that a neighbor information element that is less frequently occurring more strongly causes the target context vector to be adjusted to be closer to that neighbor context vector than a neighbor information element that is more frequently occurring.

12. The method of claim 1, further comprising:

for at least one record comprising a plurality of information elements, determining a summary context vector for the record from the normalized sum of the context vectors of the information elements that comprise the record.

13. The method of claim 12, further comprising:

receiving a query comprising at least one information element;

generating a query context vector from the information element that comprise the query;

retrieving at least one record having a summary context vector with a orientation in the vector space similar to the orientation of the query context vector.

14. The method of claim 13, wherein retrieving at least one record having a summary context vector with a orientation in the vector space similar to the orientation of the query context vector further comprising:

a tree walk of a cluster tree, the cluster tree comprising a hierarchical plurality of nodes, each node having a cluster centroid vector, each cluster centroid vector associated with a cluster of one or more records, and derived from the one or more records contained in the cluster,

the tree walk performed by iternatively selecting a node of the cluster tree that has a centroid cluster vector with a closest orientation in the vector space to the query context vector.

15. The method of claim 1, further comprising:

clustering the context vectors into a plurality of clusters, each cluster having a centroid vector derived from the plurality of context vector contained in the cluster.

16. The method of claim 1, wherein:

information elements having similar meaning have context vectors with similar orientations in the vector space.

17. The method of claim 1, further comprising:

determining a similarity of meaning between a first information element and a second information by performing a vector operation on the context vectors of the first and second information elements.

18. A computer implemented method of generating vectors representing information items for retrieval of the information elements, the method comprising:

selecting a set of R information elements;

determining for the selected set of information elements an RxR mutual co-occurrence matrix based on proximal co-occurences of the information elements in a plurality of documents;

applying Singular Value Decomposition to the mutual co-occurrence matrix to produce a set of first context vectors, the first context vectors having orientations in a D dimensional vector space, where D<<R; and

wherein each first context vector is uniquely associated with one of the selected information elements, and wherein information elements having similar meaning have respective first context vectors with similar orientations in the vector space.

19. The method of claim 13, wherein the mutual co-occurrence matrix comprises for each pair of selected information elements, a normalized measure of the frequency of proximal co-occurrence of the pair of selected information elements.

20. The method of claim 13, wherein the selected set of information elements comprises a first selected set, the method further comprising:

selecting a second set of information elements different from the first selected set of information elements;

associating each of the second set of information elements with a second context vector; and

updating the second context vectors at least once using the first context vectors, wherein the first context vectors are fixed during at least one update of the second context vectors.

21. A computer implemented method of retrieving a record from a database containing a plurality of records, each record containing at least one information element having an associated context vector, the method comprising:

for each of a plurality of information elements, storing a context vectors uniquely associated with the information element, the context vectors having the properties that information elements having similar meaning have context vectors with similar orientations in a vector space, and information elements having dissimilar meanings have context vectors with dissimilar orientations in the vector space;

for each of the plurality of records, storing a summary context vector derived from context vectors respectively associated with information elements that comprise the record;

receiving a query;

deriving at least one query information element from the query;

generating a query context vector from the query information element; and

selecting at least one record having a summary context vector with orientation in the vector space that is similar to the orientation of the query context vector.

22. The method of claim 21, wherein selecting at least one record further comprises:

for each of a plurality of records, determining a distance in the vector space between the query context vector and a summary context vector of a record; and

selecting the record having a least distance between its summary context vector and the query context vector.

23. A computer implemented of providing a universal meaning space for human understandable information elements, the method comprising:

selecting a set of first information elements;

creating a first set of context vectors based on proximal co-occurrences of the first information elements in corpus of records, each first context vector uniquely associated with one of the first information elements, the context vectors having an orientation in a vector space, such that first information elements having similar meaning have context vectors with similar orientations in the vector space;

selecting a set of second information elements, the second information elements different from the first information elements;

selecting a subset of the first information elements;

for each first information element in the subset, selecting a corresponding second information element having a human understandable meaning substantially identical to the meaning of the first information element;

for each of the selected second information elements, associating the second information element with the context vector of the corresponding first information element;

assigning a context vector to each non-selected second information element; and

adjusting the context vectors of the non-selected second information elements using the context vectors of the selected second information elements.

24. The method of claim 23, wherein:

the first information elements are words of a first human language, and the second information elements are words of a second, different human language; and

the subset of first information elements and the corresponding second information elements have substantially identical meaning.

25. The method of claim 23, wherein:

the first information elements are symbolic representations of words of a human language encoded in a first data format, and the second information elements are symbolic representations of non-text data encoded in a second data format different from the first format; and

the subset of first information elements and the corresponding second information elements have substantially related meaning even though they have different data formats and different symbolic representations.

26. The method of claim 23, wherein adjusting the context vectors of the non-selected second information elements further comprises:

adjusting the context vectors of the non-selected second information elements, such that non-selected second information elements and selected second information elements having similar meaning have context vectors with similar orientations in the vector space.

27. A computer-implemented process of generating a dictionary of information elements for a database of records, each record including at least one information element, each information element associated with a context vector, each information element having a determinate proximity to other information elements in a record, wherein a neighbor information element is an information element that occurs proximate a target information element in at least one record in the database, the method comprising:

initializing the context vectors associated with information elements in the dictionary, such that initial context vectors are substantially orthogonal to each other in a vector space;

for each information element being a target information element:

selecting neighbor information elements of the target element in at least one record;

modifying the context vector of the target information element using the context vectors of each selected neighbor information elements as a function of the proximity of each neighbor information element to the target information element, and a co-importance of the target information element and the neighbor information element.

28. The method of claim 27, further comprising:

determining the co-importance according to the relative importance of the target information element and the relative importance of the neighbor information element.

29. The method of claim 28, wherein determining the co-importance comprises:

determining a first relative importance of the target information element, inversely according to the frequency of occurrence of the target information element in the records;

determining a second relative importance of the neighbor information element inversely according to the frequency of occurrence of the neighbor information element in records; and

determining the co-importance as a function of the first relative importance and the second relative importance.

30. The method of claim 29, further comprising:

determining the relative importance of an information element by the equation:

$$I_j = B + (1 - B)(1 - \frac{\log(\frac{1}{ND_j})}{\log(\frac{1}{TND})}),$$

wherein:

I_j is the relative importance of information element J;

B represents a predefined lower bound;

ND_j represents the number of records containing information element J; and

TND represents the total number of records.

31. The method of claim 27, wherein the co-importance of the target information element and the neighbor information element is determined using the equations:

1) $C_{TN} = I_T I_N$

wherein:

C_{TN} is the co-importance of the target information element and the neighbor information element;

I_T is the relative importance of the target information element; and

I_N is the relative importance of the neighbor information element; and

2) $I_j = B + (1 - B)(1 - \frac{\log(\frac{1}{ND_j})}{\log(\frac{1}{TND})})$,

wherein:

I_j is the relative importance of an information element J;

B represents a predefined lower bound;

ND_j represents the number of records containing information element J; and

TND represents the total number of records.

32. The method of claim 27, wherein a proximity constraint varies a magnitude of the modification to the context vector of the target information element as a function of both the frequency of occurrence of the target information element and the frequency of the occurrence of each neighbor information element in the records, so that the context vectors of information elements that frequently proximally co-occur do not converge.

33. In a computer system including a storage device containing a plurality of records, each record containing a plurality of information elements, a computer readable medium for configuring and controlling the computer system to generate a plurality of context vectors, the computer readable medium comprising:

an initial context vector generation module, adapted to read and write to the storage device, which initializes to each of a plurality of selected information element an initial context vector, such that the initial context vectors are substantially orthogonal to each other in a vector space, and which writes the initial context vectors to the storage device in association with respective information elements;

a vector training module, adapted to read and write to the storage device, for modifying the context vector of a selected information element, being a target information element, using the context vectors of neighbor information elements that proximally co-occur with the target information element, as a function of the proximity of each neighbor information element to the target information element, and a co-importance of the target information element and the neighbor information element.

34. A method of automatically indexing documents using a defined index of terms, the method comprising:

providing an indexed collection of documents, each document having at least one index term assigned to the document;

providing a plurality terms, including the index terms, each term associated with a context vector, the context vector having the properties that that terms having similar meaning have context vectors with similar orientations in a vector space, terms having dissimilar meanings have context vectors with dissimilar orientations in the vector space, and terms which frequently proximally co-occur have context vectors with similar orientations in the vector space; and

generating for each indexed document a context vector from the context vectors of selected terms that comprise the document;

receiving a new document to be indexed;

generating a new context vector of the new document, the new context vector generated from the context vectors of selected terms that comprise the new document;

selecting at least one indexed document having a context vector similar to the new context vector;

assigning to the new document at least one index term assigned to a selected indexed document.

35. The method of claim 34, wherein assigning to the new document at least one index term further comprises:

for each selected indexed document, assigning a weight to each index term assigned to the indexed document, the weight proportional to the similarity between the new context vector and the context vector of the indexed document, such that the weight is higher where the context vectors are more similar;

for each index term, generating an index term score as a function of a number of occurrences of the index term in each selected indexed document, and the weight of the index term with respect to each selected indexed document, such that the index term score is higher as the number of occurrences of an index term increases; and

assigning to the new document at least one of the index terms with a high index term score.